Amendments to the Drawings

Attached is a replacement drawing of Fig. 1, corrected for errors of a clerical nature.

Attachment: Annotated marked-up drawing

Replacement drawing

REMARKS

With entry of the foregoing amendment, Claims 1-18 are pending in the application.

Claims 6 and 7 have been amended to correct misspellings of the word "grammar." Applicants respectfully request that this amendment be entered.

Figure 1 has been amended to correct errors in block 15. The text, "T6 = 'PAGE CONTINES HEADER STARTING WITH WORD 'ABOUT..'" has been replaced with the following text: "T6 = 'PAGE CONTAINS HEADER STARTING WITH WORD 'ABOUT'' " (misspelling of the word "contains" corrected and periods after the word "about" deleted).

Applicants thank the Examiner for taking the time to discuss the Final Office Action in a telephone interview. Hereinbelow Applicants present claim amendments and arguments addressing issues raised in the interview.

Claims 1-3, 6, 7, 10-12, and 15-17 have been rejected under 35 U.S.C. 102(e) as being anticipated by Russell-Falla et al. (U.S. Pat. No. 6,675,162). Claims 8, 9, and 18 have been rejected under 35 U.S.C. 103(a) as being unpatentable over Russell-Falla et al. Claims 4, 5, 13, and 14 have been rejected under 35 U.S.C. 103(a) as being unpatentable over Russell-Falla et al. in view of Haug et al. (U.S. Pat. No. 6,556,964). The rejections are respectfully traversed and reconsideration is requested.

The present invention relates to a computer method and apparatus that determines the content type of a subject Web page. In one embodiment, a predefined set of potential content types is first provided as part of a preparation phase. Then, a distinguishing series of tests is prepared for each content type to be identified (See Specification page 9, lines 7-11, 24-26, and page 14, lines 13-24). For each potential content type, the distinguishing series of tests, when applied, have test results which enable a quantitative evaluation of the contents of the subject Web page. The distinguishing series of tests may include: (1) determining whether a predefined piece of data or keyword appears in the subject Web page, (2) examining syntax or grammar or text properties, (3) examining page format and style, (4) examining links in the subject Web page, and (5) examining links that refer to the subject Web page (see Specification page 11, line 20 to page 12, line 8 and Claims 6, 7, 16, and 17).

Examples of tests that examine syntax or grammar or text properties include the number of passive sentences, the number of sentences without a verb, and the percentage of verbs in the past tense. (see Specification page 12, lines 1-2). Examples of tests that specifically examine text properties include "[t]he first sentence or the first paragraph has a date" and "[t]he average sentence length in the page is in one of the following ranges: 1=[0-65), 2=[65-∞]" (see Specification page 19, lines 9 and page 18, lines 21-22).

Given a subject web page, each of the distinguishing series of tests is applied. Based on the test results, a respective probability for each potential content type being detected in some contents of the subject Web page is determined. A series of Bayesian networks or other means is used to combine the probabilities from the test results to provide indications of the types of contents detected in the subject Web page.

The Russell-Falla et al. patent describes computer-implemented methods for characterizing a specific category of information content within certain types of media such as "web pages, e-mail, and other types of digital datasets" (col. 1, lines 15-16). With regard to Web pages, the characterization occurs by examining the content of a particular Web page and determining whether to display the Web page contents to the user. The examining step "includes identifying and analyzing the web page natural language content relative to a predetermined database of weighted words - or more broadly regular expressions - to form a rating" (col. 2, lines 55-59).

The database of weighted expressions, also called the "Target Attribute Set," is created according to a process described in Russell-Falla et al. under the heading "Formulating Weighted Lists of Words and Phrases" at col. 6, line 22. As shown in Figure 3 of Russell-Falla et al., training pages 82, including "Good" web pages 84 and "Bad" web pages 86, are provided. Next, a list of unique words and expressions is created for each of the training pages 82. Data indicating the frequency at which each expression in the training pages 82 occurs is statistically analyzed 90 to produce the Target Attribute Set 92. The Target Attribute Set contains attributes or expressions indicating a particular type of content. The attributes or expressions are also input to a Correlation Engine which "searches for correlations between attributes across content sets." Expressions that "appear frequently in both sets without mitigating correlation are discarded. The remaining attributes constitute the Target Attribute Set." Weights are then assigned to the

expressions using a neurol-network/back-propagation technique. After this preparation and training process, the Target Attribute Set may be used to classify any given web page.

Thus, Russell-Falla et al. performs only one type of test in which certain words or phrases within a subject Web page are compared with a list of words or phrases associated with a particular category of information content.

Base Claims 1 and 10 have been rejected under 35 U.S.C. 102(e) as being anticipated by Russell-Falla et al. because Russell-Falla et al. is purported to teach "preparing a distinguishing series of tests ... (i.e., testing each keyword or regular expression of the content against a database of keywords and regular expressions common to the content type for matching and weighting purposes)" (Final Office Action, page 3, Section 6). The one and only type of test taught by Russell-Falla et al. is encompassed by the test "determining whether a predefined piece of data. keyword, or expression appears in the subject Web page" that is one of a series of exemplary test types recited in Claims 6, 7, 16, and 17 of the present Application ("data" includes "digital information" referred to at col. 6, line 35). Indeed, Russell-Falla et al. teach "running this test a plurality of times for different predetermined pieces of data for a specific content type" (Final Office Action, page 6, Section 9). Russell-Falla et al. do not teach, however, other types of tests besides the test that determines whether predefined keywords or regular expressions appear in any given Web page. The correlation engine of Russell-Falla et al. is not a separate test, but is part of the process of creating the one and only test described by Russell-Falla et al. (i.e., formulating weighted lists of expressions) (see col. 6, line 52 through col. 7, line 24). In other words, the correlation engine is used to help prepare the Target Attribute Sets or weighted lists of expressions that are used to classify a given web page. Therefore, Applicants have amended Claims 1 and 10 to make clear that at least one of the plural distinguishing series of tests does not determine whether a predetermined piece of data, keyword or expression appears in the subject Web page.

Because Russell-Falla et al. do not teach all of the elements within amended base Claims 1 and 10 (". . . prepare a distinguishing series of tests, wherein at least one test does not determine whether a predetermined piece of data, keyword or expression appears in the subject Web page"), the Applicant respectfully submits that the § 102 rejection regarding independent Claims 1 and 10 should be withdrawn.

Because Claims 2, 3, 6, 7, 11, 12, and 15-17 depend from and are limited by base Claims 1 and 10 respectively, the § 102 rejection of these claims should be withdrawn for at least the same reasons.

Because there is no suggestion or motivation within Russell-Falla et al. to use a distinguishing series of tests, wherein at least one test does not determine whether a predetermined piece of data or keyword appears in the subject Web page, the Final Office Action fails to make a case of prima facie obviousness regarding now amended base Claims 1 and 10. Claims 8, 9, and 18 depend from and are limited by base Claims 1 and 10 respectively, which are argued above to be in an allowable condition. Accordingly, the § 103 rejection of Claims 8, 9, and 18 based on Russell-Falla et al. should be withdrawn.

With regard to Claims 4, 5, 13, and 14 being rejected under 35 U.S.C. 103(a) as being unpatentable over Russell-Falla et al. in view of Haug et al., Haug et al. do not add the use of a distinguishing series of tests, wherein at least one test does not determine whether a predetermined piece of data or keyword appears in the subject Web page, which is lacking from Russell-Falla et al. Because the combination of Russell-Falla et al. and Haug et al. do not make a case of prima facie obviousness regarding the invention as now claimed in base Claims 1 and 10, the § 103 rejection of Claims 4 and 5, which are dependent from and limited by base Claim 1, should be withdrawn. Also, the § 103 rejection of dependent Claims 13 and 14, which are dependent from and limited by base Claim 10, should be withdrawn.

CONCLUSION

In view of the above amendments and remarks, it is believed that all claims (Claims 1-18) are in condition for allowance, and it is respectfully requested that the application be passed to issue. If the Examiner feels that a telephone conference would expedite prosecution of this case, the Examiner is invited to call the undersigned.

Respectfully submitted,

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

By _____

Mary Lou Wakimura
Registration No. 31,804
Telephone: (978) 341-0036
Facsimile: (978) 341-0136

Concord, MA 01742-9133
Dated: 5/13/05

Appl'n No.: 09/768869
Title: Computer Method and Apparatus. . .
Inventors: Kosmas Karadimitriou et al.
Annotated Sheet

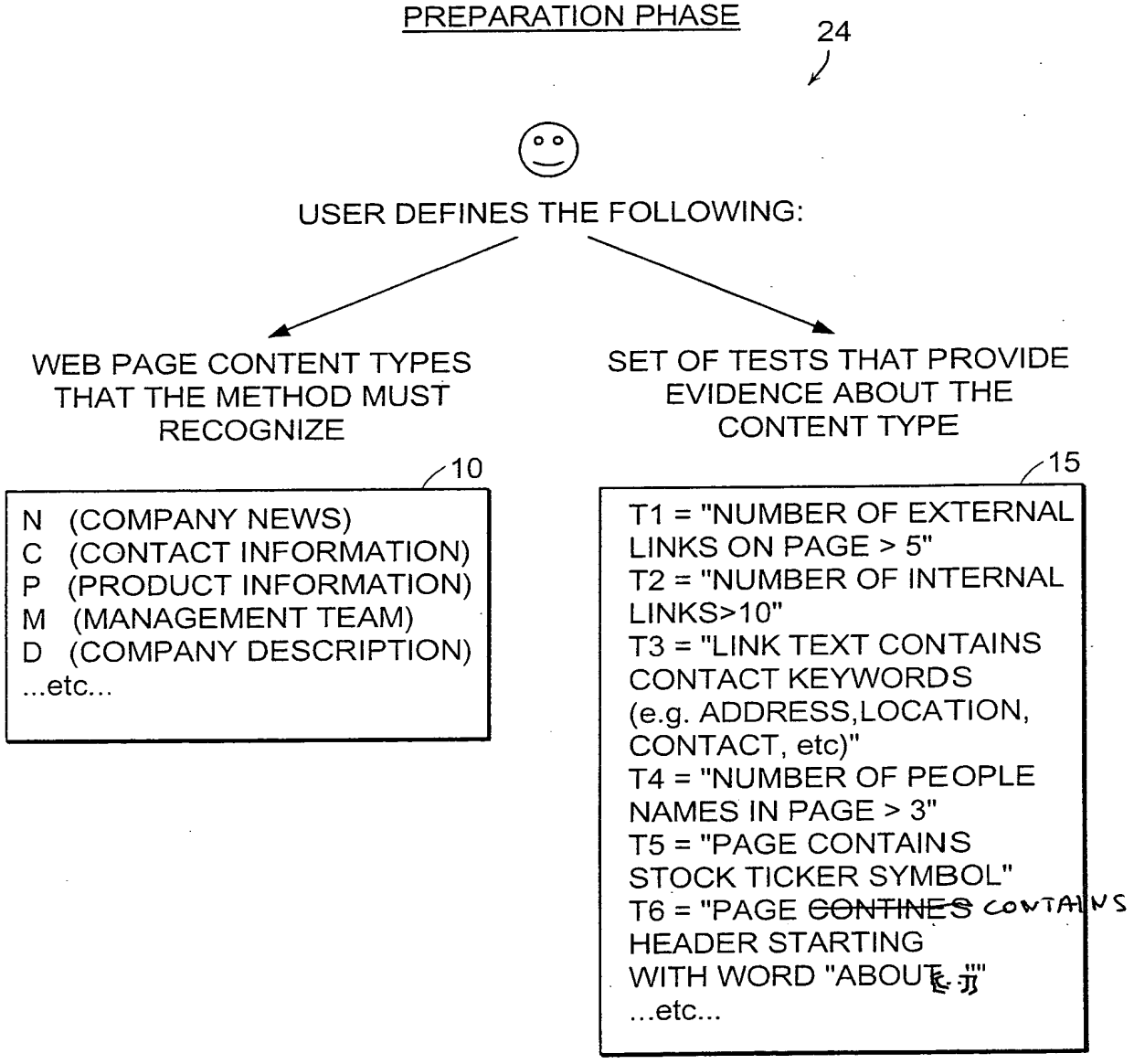

FIG. 1